How Can Machine Learning Increase Students' exposure to Career Opportunities?

Use of NLP to increase exposure to Career Opportunities

Joshua Kiprono

06 April 2022

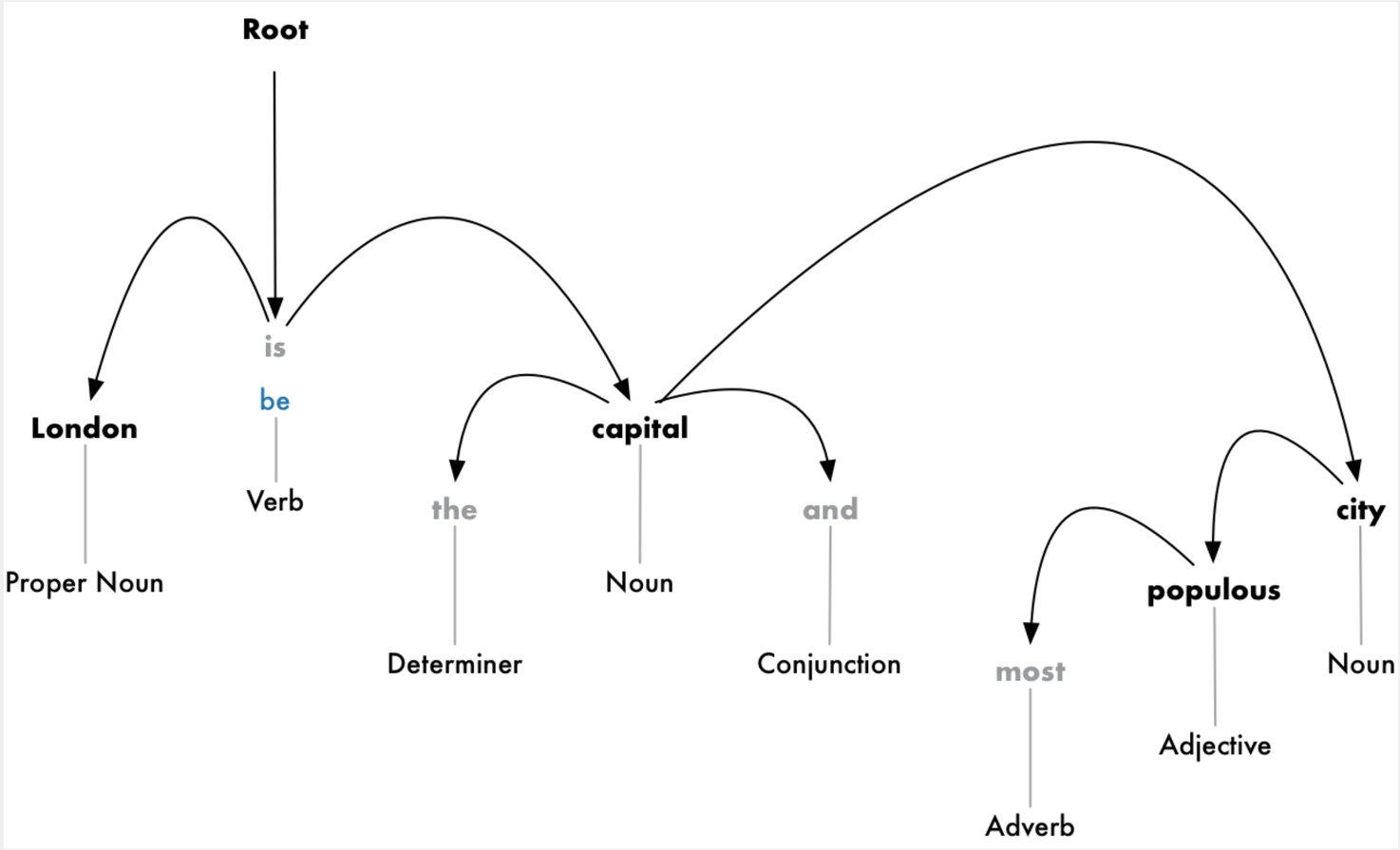
What is Natural Language Processing

NLP is a subfield of linguistics, computer science, and artificial intelligence concerned with the interactions between computers and human language, in particular how to program computers to process and analyze large amounts of natural language data.

The goal of Natural Language Processing

The goal is a computer capable of "understanding" the contents of documents, including the contextual nuances of the language within them. The technology can then accurately extract information and insights contained in the documents as well as categorize and organize the documents themselves.

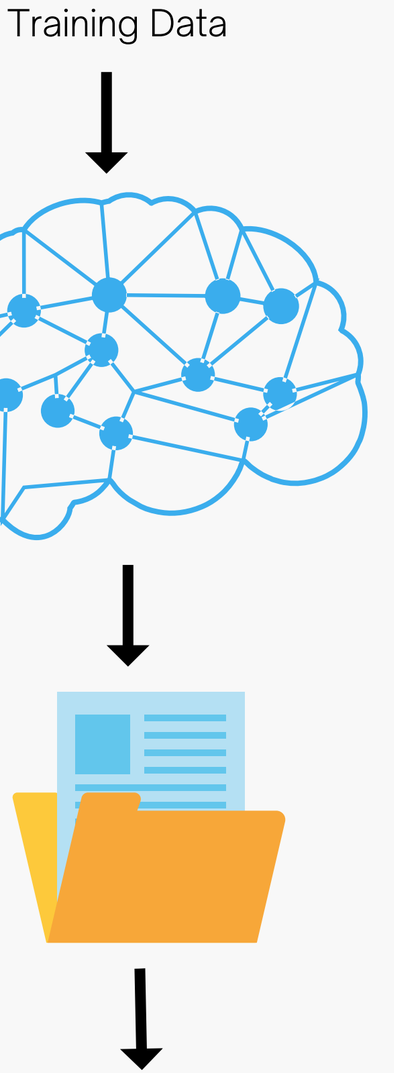
How NLP works



Building a pipeline

The idea is to break up your problem into very small pieces and then use machine learning to solve each smaller piece separately.

Machine Learning Algorithms and NLP



Problem Statement

The goal of Office of Career Development is to help as many students as possible achieve their career objectives yet there is no established system to help the office make decisions based on real time data.

Note that a 2016 Gallup poll showed that 61 percent of students who graduated between 2010 and 2016 said they had visited a career center at least once and 17 percent found it *very effective*

In this research, I will be investigating how to use machine learning to increase the effectiveness of the Office of Career Development in sending career opportunities and for students to increase their productivity while finding career opportunities.

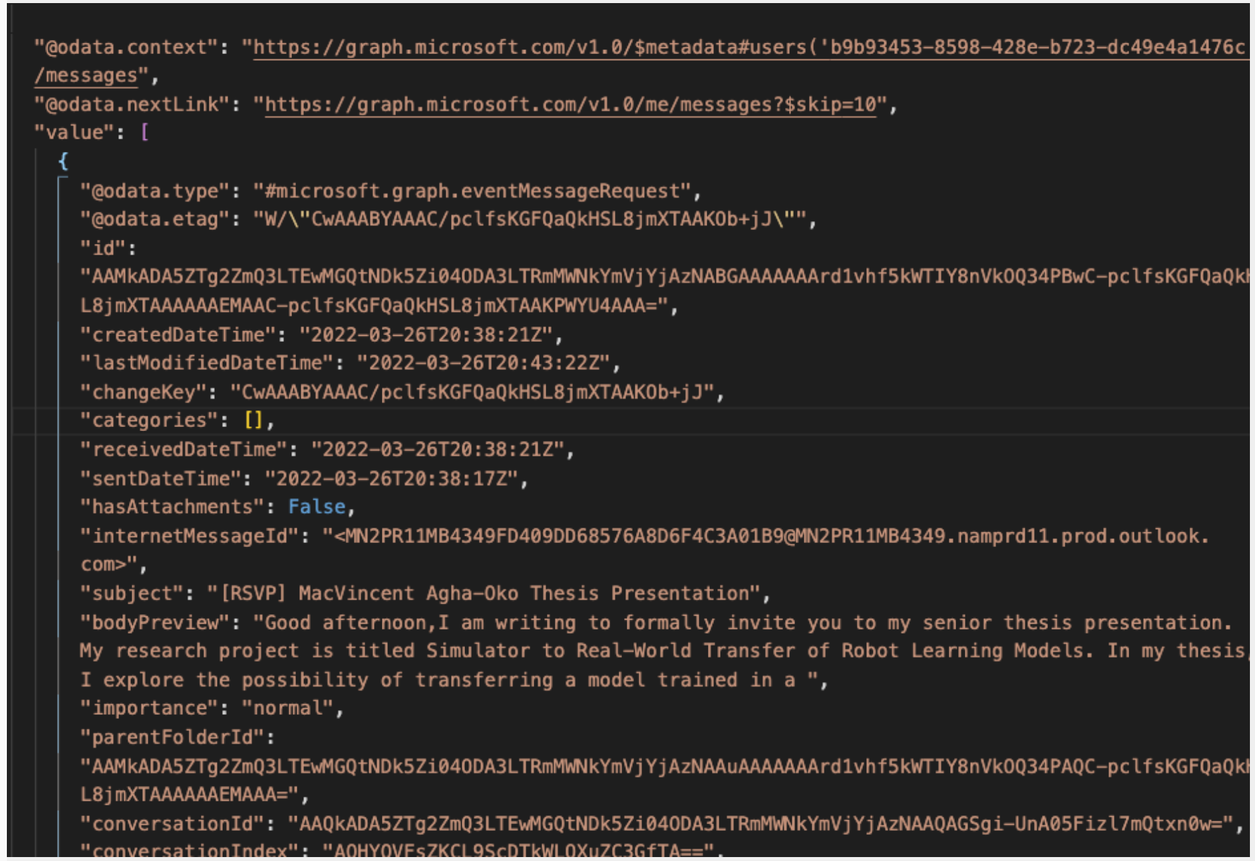
Methodology

- Using an experimental approach
- Have a set of students using an application and others without the application and find out the difference in outreach of such career opportunities

1. Creating a test application

Application Framework: Django
Reason: Lightweight and fast
Purpose: Test keyword extraction and rule-based matching
Requirements: Spacy(free open-source library for Natural Language Processing in Python), Microsoft Graph

2. Retrieving Emails



- User gives permission for the application to access emails through graph
- Emails are received in JSON format with various other datapoints

2. Cleaning data

- process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset

```
args = [
                ("\\\r\\\n", " "), #Removes \r\n
                ("(http|ftp|https)://([\w_-]+(?:(?:\.[\w_-]+)+))([\w.,@?^=%&:/~+#-]*[\w@?^=%&/~+#-])?", " "), #
                ("\\< \\>", " "), #removes < >
                ("\\[ \\]", " "),# Removes [\\]
                ("\\[.*?\\]", " "), #Removes text within [ ]
                ("\\<.*?\\>", " "), #Removes text within <>
                ("_____________________________", " "), #Removes dashes
                ("\\*", " "), #Removes dashes
                ("CAUTION: This email originated from outside the organization. Do not click links or open atta
                ("  ", " "),
                ("[ ]{2,}"," "), #Removes every empty space whose len is greater than 1
    ]
def clean_data(args, text):
    # Create a regular expression  from the tuple
    for old, new in args:
        text = re.sub(old, new, text)
    return text
```

Token-based matching

- spaCy features a rule-matching engine, the Matcher, that operates over tokens, similar to regular expressions

```
def claflin():
    matcher = Matcher(nlp.vocab)
    #Needs improvement
    pattern = [
            [{"LOWER": "apply"}],[{"LOWER": "opportunity"}],[{"LOWER": "internship"}],[{"LOWER": "intern"}], [{
            ]
    finalData = {"mimetype":"application/json"
    } #Holds cleaned data whose senders are from Claflin and whose body contains the above keywords
    arr = []
    for mail in data:
        subdata = {}
        claflinMail = re.search('@claflin\\.edu$', mail['sender']['emailAddress']['address'] )
        if claflinMail:
            #Save links in a separate map for later use
            subdata['Links'] = re.search("(http|ftp|https)://([\w_-]+(?:(?:\.[\w_-]+)+))([\w.,@?^=%&:/~+#-]*[\w
            body = clean_data(args, mail['body']['content'])
            doc = nlp(body)
            #text1= spacy.NER(bod)
            entities = [(ent.text, ent.label_) for ent in doc.ents]
            matcher.add("OpportunityKeys", pattern)
            matches = matcher(doc)
            for match_id in matches:
                if match_id and (mail['id'] not in finalData):
                    document = nlp(body)
                    entities = [(ent.text, ent.label_) for ent in document.ents]
                    finalData[str(mail["id"])] = {"body": body, "subject":mail["subject"], "sender": mail["send
    return finalData
```

Results

2022 @ 4:00pm Resident Assistants (RAs) are undergraduate students who live and work in Claflin University residence halls. RAs are assigned to a particular wing or floor in each residence hall and are responsible for fostering community among fellow resident students by developing social and educational opportunities, assisting students in personal and administrative matters, and ensuring that the standards set forth in the student code of conduct are consistently being met. The RA position provides students with many opportunities to develop leadership skills, plan programs and events, exercise peer advising and mediating, as well as important administrative skills that will help in any career down the road. Serving as an RA is a one-year commitment that may be renewed at the discretion of the Executive Director of Residential Life and Housing. Full-time rising sophomores, juniors, and seniors \u2014 especially student leaders with involvement in artistic, athletic or ROTC programs \u2014 are encouraged to apply. Current commuters are encouraged to apply as well who decide to reside on campus. Mr. Dillon Beckford- US Army Veteran | M.Ed Claflin University \u2013 Student Development and Services Assistant Dean, Executive Director of Housing and Residential Life and Student Conduct 400 Magnolia St., Orangeburg, SC 29115 Phone: 803-535-5301 | Fax: 803-535-5618 |Email: dbeckford@claflin.edu ", "entities": [["&", "ORG"], ["&", "ORG"], ["RA APPLICATION", "ORG"], ["March 4, 2022 @ 4:00pm Resident", "DATE"], ["Claflin University", "ORG"], ["RA", "ORG"], ["RA", "ORG"], ["one-year", "DATE"], ["Residential Life", "ORG"], ["Dillon", "PERSON"], ["US Army", "ORG"], ["Claflin University \u2013 Student Development and Services", "ORG"], ["Housing and Residential Life and Student Conduct", "ORG"], ["400", "CARDINAL"], ["Orangeburg", "GPE"], ["803", "CARDINAL"], ["803-535-5618", "CARDINAL"], ["dbeckford@claflin.edu &", "ORG"]], "sender": { "emailAddress": { "address": "Communications@claflin.edu", "name": "Communication & Marketing" } }, "subject": "Help Wanted: Resident Assistants" },
"AAMkADA5ZTg2ZmQ3LTEwMGQtNDk5Zi04ODA3LTRmMWNkYmVjYjAzNABGAAAAAAAd1vhf5kWTIY8nVkOQ34PBwC-pclfsKGFQaQkHSL8jmXTAAAAAAEMAAC-pclfsKGFQaQkHSL8jmXTAAJ06RJ2AAA=": { "body": " Apply Now for the First Lady\u2019s Hats & Gloves Tea Endowed Scholarship DEADLINE TO SUBMIT: FEBRUARY 25, 2022 Submit completed application (attached) and supporting documentation to : Co-chairs of the Scholarship Committee: Dr. Peggy S. Ratliff pratliff@claflin.edu (803) 535-5233 Mrs Marilyn G. Drayton mgibbsdrayton@claflin.edu (803) 535-5309 Claflin University 400 Magnolia Street Orangeburg, S.C. 29115 ", "entities": [["Apply Now", "ORG"], ["First", "ORDINAL"], ["Gloves Tea Endowed Scholarship &", "ORG"], ["&", "ORG"], ["DEADLINE TO SUBMIT", "ORG"], ["FEBRUARY 25, 2022 &", "DATE"], ["&", "ORG"], ["Submit", "ORG"], ["&", "ORG"], ["the Scholarship Committee: &", "ORG"], ["&", "ORG"], ["Peggy S. Ratliff ", "PERSON"], [" ", "ORG"], ["pratliff@claflin.edu ", "ORG"], ["803", "CARDINAL"], ["535", "CARDINAL"], ["Marilyn G.", "PERSON"], ["mgibbsdrayton@claflin.edu", "ORG"], ["803", "CARDINAL"], ["535-5309 &", "PERCENT"], ["Claflin University", "ORG"], ["Orangeburg", "GPE"], ["S.C.", "GPE"], ["29115", "CARDINAL"], ["&", "ORG"], ["&", "ORG"], ["&", "ORG"], ["&", "ORG"], ["&", "ORG"], ["&", "ORG"]], "sender": { "emailAddress": { "address": "Communications@claflin.edu", "name": "Communication & Marketing" } }, "subject": "Apply Now for the First Lady\u2019s Hats & Gloves Tea Endowed Scholarship" },
"AAMkADA5ZTg2ZmQ3LTEwMGQtNDk5Zi04ODA3LTRmMWNkYmVjYjAzNABGAAAAAAAd1vhf5kWTIY8nVkOQ34PBwC-pclfsKGFQaQkHSL8jmXTAAAAAAEMAAC-pclfsKGFQaQkHSL8jmXTAAJ06RJ6AAA=": { "body": " Please take advantage of this awesome opportunity. Apply Now for the First Lady\u2019s Hats & Gloves Tea Endowed Scholarship DEADLINE TO SUBMIT: FEBRUARY 25, 2022 Submit completed application (attached) and supporting documentation to : Co-chairs of the Scholarship Committee: Dr. Peggy S. Ratliff pratliff@claflin.edu (803) 535-5233 Mrs. Marilyn G. Drayton mgibbsdrayton@claflin.edu (803) 535-5309 Claflin University 400 Magnolia Street Orangeburg, S.C. 29115 ", "entities": [["Apply Now", "ORG"], ["First", "ORDINAL"], ["Gloves Tea Endowed Scholarship &", "ORG"], ["&", "ORG"], ["DEADLINE TO SUBMIT", "ORG"], ["FEBRUARY 25, 2022 &", "DATE"], ["&", "ORG"], ["Submit", "ORG"], ["&", "ORG"], ["the Scholarship Committee: &", "ORG"], ["&", "ORG"], ["Peggy S. Ratliff ", "PERSON"], [" ", "ORG"], ["pratliff@claflin.edu ", "ORG"], ["803", "CARDINAL"], ["535", "CARDINAL"], ["Marilyn G.", "PERSON"], ["mgibbsdrayton@claflin.edu", "ORG"], ["803", "CARDINAL"], ["535-5309 &", "PERCENT"], ["Claflin University", "ORG"], ["Orangeburg", "GPE"], ["S.C.", "GPE"], ["29115", "CARDINAL"], ["&", "ORG"], ["&", "ORG"], ["&", "ORG"], ["&", "ORG"], ["&", "ORG"], ["&", "ORG"]], "sender": { "emailAddress": { "address": "bebrown@claflin.edu", "name": "Bernay M. Brown" } }, "subject": "FW: Apply Now for the First Lady\u2019s Hats & Gloves Tea Endowed Scholarship" }, "AAMkADA5ZTg2ZmQ3LTEwMGQtNDk5Zi04ODA3LTRmMWNkYmVjYjAzNABGAAAAAAAd1vhf5kWTIY8nVkOQ34PBwC-pclfsKGFQaQkHSL8jmXTAAAAAAEMAAC-pclfsKGFQaQkHSL8jmXTAAJ06RJNAAA=": { "body": " Good morning students, Hope you are well today, the information below is for your review. Please read over the information below and respond as needed. Have a great day. Sherry Bridges Administrative Assistant for Mathematics & Computer Science School of Natural Sciences & Mathematics James S. Thomas Science Building, Room 125 400 Magnolia Street Orangeburg, South Carolina 29115 Email - shbridges@claflin.edu From: Jones, Paula M. < jonespa@musc.edu > Sent: Saturday, February 12, 2022 2:19 PM To: jray@allenuniversity.edu < jray@allenuniversity.edu >; Sonya Johnson < Sonya.Johnson@benedict.edu >; etosha.johnson@benedict.edu < etosha.johnson@benedict.edu >; Pamela Shuler < pshuler@claflin.edu >; asharper@clintoncollege.edu < asharper@clintoncollege.edu >; flin@coastal.edu < flin@coastal.edu >; shuklam@denmarktech.edu < shuklam@denmarktech.edu >; thicks@morris.edu < thicks@morris.edu >; tharding@scsu.edu < tharding@scsu.edu >; stukes@scsu.edu < stukes@scsu.edu >; tgjackson@voorhees.edu < tgjackson@voorhees.edu >; zgolkar@voorhees.edu < zgolkar@voorhees.edu >; graggm@atc.edu < graggm@atc.edu >; smims@csuniv.edu < smims@csuniv.edu >; harolyw@clemson.edu < harolyw@clemson.edu >; eippertk@cofc.edu < eippertk@cofc.edu >; gadsonsc@octech.edu < gadsonsc@octech.edu >; sjohnson@swu.edu < sjohnson@swu.edu >; pcapers@citadel.edu < pcapers@citadel.edu >; AR60@uscb.edu < AR60@uscb.edu >; dimaculangad@winthrop.edu < dimaculangad@winthrop.edu >; allisonjm@cofc.edu < allisonjm@cofc.edu >; ppeart@presby.edu < ppeart@presby.edu >; Leigh Lanford < lwlanford@presby.edu > Subject: SC AHEC Bench to Bedside Educational Session - February 16, 2022 Good afternoon, Our next educational seminar will take place on Wednesday, February 16, 2022 at 5:00 pm via our WebEx platform. Please share the below information with students. February 16 B2B: Financial Aid \u2013 The Best Bang for Your Buck! Presenter: Rodney Jones Financial Aid Counselor Office of Enrollment Services Medical University of South Carolina Description: Financial aid is monetary assistance that can help you pay for college. Some aid needs to be paid back or earned, and some aid is a gift. This session will provide you with important information you should know about financial aid. Objectives: Upon completion of this seminar, students should be able to: 1) Identify the different types of financial aid 2) Understand the Free Application for Federal Student Aid (FAFSA) process 3) Identify service and scholarship opportunities 4) Understand Cost of Attendance (COA) and creating a personal budget REGISTRATION LINK B2B: Financial Aid \u2013 The Best Bang for Your Buck! If you have questions, please contact Ms. Jones. Paula M. Jones, MBA, BHS Director, Health Careers Program South Carolina Area Health Education Consortium 1 South Park Circle Suite 203 Charleston, SC 29407 O: 843-792-4443 F: 843-792-4430 WWW.SCAHEC.NET CAUTION: This email originated from outside the organization. Do not click links or open attachments unless you recognize the sender and know the content is safe. ", "entities": [["morning", "TIME"], ["today", "DATE"], ["Sherry Bridges Administrative", "PERSON"], ["Mathematics &", "ORG"], ["Computer Science School of Natural Sciences &", "ORG"], ["James S. Thomas Science Building", "PERSON"], ["400", "CARDINAL"], ["Magnolia Street Orangeburg", "FAC"], ["South Carolina", "GPE"], ["29115", "DATE"], ["Email - &", "ORG"], ["Paula M. &", "PERSON"], ["jonespa@musc.edu >", "ORG"], ["Saturday, February 12, 2022", "DATE"], ["2:19 PM", "TIME"], ["jray@allenuniversity.edu <", "ORG"], ["jray@allenuniversity.edu >", "ORG"], ["Sonya Johnson &", "ORG"], ["Sonya.Johnson@benedict.edu >", "ORG"], ["etosha.johnson@benedict.edu <", "ORG"], ["etosha.johnson@benedict.edu >", "ORG"], ["Pamela Shuler &", "ORG"], ["pshuler@claflin.edu >", "ORG"], ["asharper@clintoncollege.edu &", "ORG"], ["asharper@clintoncollege.edu >", "ORG"], ["flin@coastal.edu &", "ORG"], ["flin@coastal.edu >", "ORG"], ["shuklam@denmarktech.edu <", "ORG"], ["shuklam@denmarktech.edu >", "ORG"], ["thicks@morris.edu <", "ORG"], ["thicks@morris.edu >", "ORG"], ["tharding@scsu.edu <", "ORG"], ["tharding@scsu.edu >", "ORG"], ["stukes@scsu.edu <", "ORG"], ["stukes@scsu.edu >", "ORG"], ["tgjackson@voorhees.edu &", "ORG"], ["tgjackson@voorhees.edu >", "ORG"], ["zgolkar@voorhees.edu &", "ORG"], ["zgolkar@voorhees.edu >", "ORG"], ["graggm@atc.edu &", "ORG"], ["graggm@atc.edu >", "ORG"], ["smims@csuniv.edu &", "ORG"], ["smims@csuniv.edu >", "ORG"], ["harolyw@clemson.edu <", "ORG"], ["harolyw@clemson.edu >", "ORG"], ["eippertk@cofc.edu &", "ORG"], ["eippertk@cofc.edu >", "ORG"], ["gadsonsc@octech.edu &", "ORG"], ["gadsonsc@octech.edu >", "ORG"], ["sjohnson@swu.edu >", "ORG"], ["pcapers@citadel.edu &", "ORG"], ["pcapers@citadel.edu >", "ORG"], ["AR60@uscb.edu &", "ORG"], ["AR60@uscb.edu >", "ORG"], ["dimaculangad@winthrop.edu <", "PERSON"], ["dimaculangad@winthrop.edu >", "ORG"], ["allisonjm@cofc.edu <", "ORG"], ["allisonjm@cofc.edu >", "ORG"], ["ppeart@presby.edu &", "ORG"], ["ppeart@presby.edu >", "ORG"], ["Leigh Lanford &", "ORG"], ["lwlanford@presby.edu >", "ORG"], ["Bedside Educational Session -", "ORG"], ["February 16, 2022 &", "DATE"], ["Good afternoon", "TIME"], ["Wednesday, February 16, 2022", "DATE"], ["5:00 pm", "TIME"], ["&", "ORG"], ["&", "ORG"], ["February 16", "DATE"], ["Buck", "PERSON"], ["Rodney Jones Financial Aid Counselor Office of Enrollment Services Medical University of South Carolina &", "ORG"], ["Description: Financial", "ORG"], ["Objectives", "ORG"], ["1", "CARDINAL"], ["2", "CARDINAL"], ["3", "CARDINAL"], ["4", "CARDINAL"], ["Attendance", "PERSON"], ["&", "ORG"], ["REGISTRATION LINK B2B", "PERSON"], ["&", "ORG"], ["Jones", "PERSON"], ["&", "ORG"], ["Paula M. Jones", "PERSON"], ["MBA", "ORG"], ["BHS Director", "ORG"], ["Health Careers Program South Carolina Area Health Education Consortium", "ORG"], ["South Park Circle", "LOC"], ["203", "CARDINAL"], ["Charleston", "GPE"], ["SC 29407", "ORG"], ["843-792-4443", "CARDINAL"], ["843-792-4430", "CARDINAL"]], "sender": { "emailAddress": { "address": "shbridges@claflin.edu", "name": "Sherry D. Bridges" } }, "subject": "Fw: SC AHEC Bench to Bedside Educational Session - February 16, 2022" },
"AAMkADA5ZTg2ZmQ3LTEwMGQtNDk5Zi04ODA3LTRmMWNkYmVjYjAzNABGAAAAAAAd1vhf5kWTIY8nVkOQ34PBwC-pclfsKGFQaQkHSL8jmXTAAJ06RJRAAA=": { "body": " Good morning Juniors, Wells Fargo is looking for talented college juniors to join our Wells Fargo Enterprise Operations Summer Internship Program. There are multiple groups the intern can join (referenced below). However, I am specifically looking for candidates that are interested in working in my group, Process Engineering and Business Process Management. The desired candidate should have a BA/BS major concentration in Industrial Engineering, Supply Chain, or Data Analytics. In addition, the candidate should have experience in process improvement and collaboration with people to understand problems and to help build solutions. A relocation allowance is provided. We are excited to announce we are currently accepting resumes for our 2022 Enterprise Operations Summer Internship Program! To be considered, express interest and submit your resume via email to Kelsey.Jensen@wellsfargo.com by February 21 2022! Internship Locations include: Charlotte, NC; St. Louis, MO; and Minneapolis, MN Seeking College Juniors to join Wells Fargo\u2019s University Program: Enterprise Operations Summer Internship Program! Now Accepting Resumes! At Wells Fargo, we are looking for talented people who will put our customers at the center of everything we do. We are seeking candidates who embrace diversity, equity, and inclusion in a workplace where everyone feels valued and inspired. Help us build a

Results and Application

- Out of the 1000 emails extracted, 153 were career related, this represents 15.3 percent of emails.
- Not found a clear metric to find the accuracy of my model

- This can be used in creating email filters to easily sift through career related opportunities.
- A full scale application can be built based on this to help instructors gauge analyze how students interact with such emails

Problems

- Email data is very sensitive and users may not oblige to share such data
- Even with users agreeing to share data, storage of such information requires high levels of data security.
- The problem can be solved without NLP, have students share opportunities instead of reliance to emails.

Solutions

- Using a smaller sample size
- Using Blockchain technology to implement the application

Near Future Goals

- Creating a full scale application
- Creating a blockchain application

References

- Marken, Z. A. and S. (2022, April 6). One in six U.S. grads say career services was very helpful. Gallup.com. Retrieved April 17, 2022, from https://news.gallup.com/poll/199307/one-six-grads-say-career-services-helpful.aspx